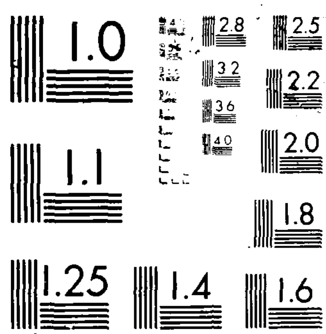
AD-A192 913 EMPIRICAL ANALYSIS AND REFINEMENT OF EXPERT SYSTEM KNOWLEDGE BASES(U) RUTGERS - THE STATE UNIV NEW BRUNSWICK NJ CENTER FOR EXPERT S.. S M WEISS 29 FEB 88 1/1

UNCLASSIFIED N00014-87-K-0398 F/G 12/9 NL


1.0
2.8
2.5
3.2
2.2
3.6
4.0
2.0
1.1
1.8
1.25
1.4
1.6

AD-A192 913

(2)

DTIC FILE COPY

DTIC
ELECTE
MAR 3 0 1988
S D
D

RUTGERS UNIVERSITY
Center for Expert Systems Research

Quarterly Report:
Empirical Analysis and Refinement of
Expert System Knowledge Bases
Contract Number
N00014-87-K-0398
Office of Naval Research

February 29, 1988

Principal Investigators:
Sholom M. Weiss
Casimir A. Kulikowski

1. Technical Project Summary

Knowledge base refinement is the modification of an existing expert system knowledge base with the goals of localizing specific weaknesses in a knowledge base and improving an expert system's performance. Systems that automate some aspects of knowledge base refinement can have a significant impact on the related problems of knowledge base acquisition, maintenance, verification, and learning from experience. The SEEK system was the first expert system framework to integrate large-scale performance information into all phases of knowledge base development and to provide automatic information about rule refinement. A recently developed successor system, SEEK2, significantly expands the scope of the original system in terms of generality and automated capabilities.

Based on promising results using the SEEK approach, ~~we~~ believe that significant progress can be made in expert system techniques for knowledge acquisition, knowledge base refinement, maintenance, and verification.

2. Principal Expected Innovations

We are proposing to demonstrate a rule refinement system in an application of the diagnosis of complex equipment failure. The expected candidate application is computer network troubleshooting. The expert system should demonstrate the following advanced capabilities:

- automatic localization of knowledge base weaknesses
- automatic repair (refinement) of poorly performing rules
- automatic verification of new knowledge base rules
- some automatic learning capabilities.

3. Objectives for FY88

- functioning equipment diagnosis and repair knowledge base, suitable for refinement (expected in the area of computer networks).
- initial demonstration of functioning equipment diagnostic system with capabilities of localization of weak rules, automatic refinement, automatic verification.
- demonstration of initial rule learning capabilities.

4. Summary of Progress

Here are the highlights of progress has been made in meeting our stated objectives for fiscal 1988:

- Dr. Peter Politakis of the Digital Equipment Co. transferred to us DEC's Network Troubleshooting Consultant program that we proposed to use in our system. Dr. Politakis directed the development of this software and will serve as our expert in the refinement of the knowledge base. We have circumscribed the knowledge base to the following problem types: line, circuit, or cable problems. This subset of the knowledge base consists of 287 observations, 138 hypotheses, and 324 rules.

- Politakis has obtained documented cases of network problems. He has supplied about a dozen so far, and we hope to obtain others from DEC's stored records. Because the case histories are stored unformatted as text, the process of extracting cases is quite tedious. We will supplement a core group of documented cases with simulated cases derived from verified correct rules in the knowledge base. (These rules may be partially hidden form the refinement system.)

- Substantial progress has been made in our rule induction (learning) system. Several experiments have been underway using data obtained from other researchers who have published results in the AI literature. These include data from Michalski [Michalski, Mozetic, Hong, and Lavrac 86] and Quinlan [Quinlan 87a, Quinlan 87b]. Our efforts are extensions of procedures we reported at the AAAI-87 conference [Weiss, Galen, and Tadepalli 87]. Additional details on the results of these are provided in the next. Complete details of the Predictive Value Optimization (PVO) procedures and results will appear soon in a technical report.

Progress in Rule Induction Techniques

Empirical techniques for induction of decision rules have evolved from procedures that cover all cases in a data base to more accurate procedures for estimating error by train and test sampling. Procedures that prune a set of decision rules and the components of these rules have been successful in increasing the performance of an induced rule set on new test cases. Recently, we reported on a technique for learning the single best decision rule of a fixed length. We have shown how jackknifing and resampling techniques for estimating error rates, can be integrated into this procedure for induction of decision rules. Superior results are reported on data sets previously analyzed in the AI literature.

In 1987, we reported on a technique for learning the *single* best decision rule of a fixed length [Weiss, Galen, and Tadepalli 87]. In contrast to other methods of rule induction, the PVO rule induction procedure does not generate and prune a complete set of decision rules. Instead, this method is an approximation to exhaustive generation of all possible rules of a fixed length. While a true exhaustive search is not feasible in most applications, a small number of heuristics reduce the search space to manageable proportions.

Figure 4-1 illustrates the key steps of the heuristic procedure.

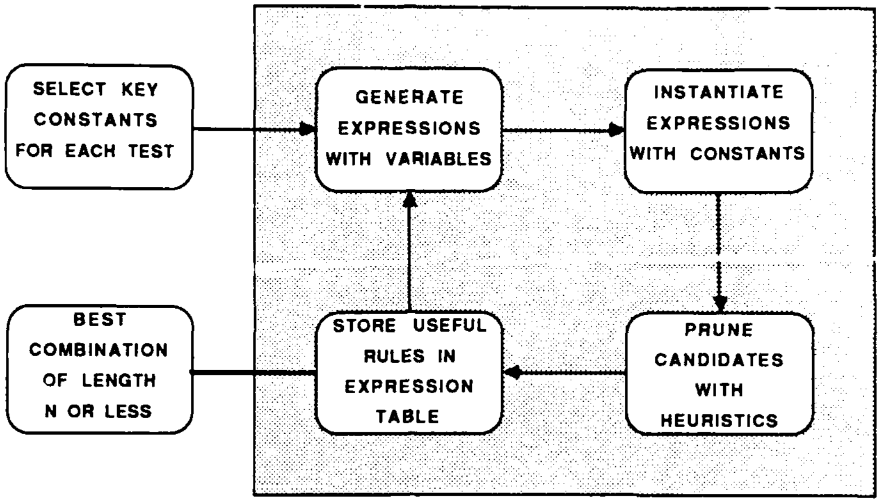


Figure 4-1: Overview of Heuristic Procedure for Best Test Combination

Experiments were performed on two sets of data for which published studies are available. The results are summarized in Figures 4-2 and 4-3.

Method	Variables	Rules	Error Rate
AQ15	7	2	32%
Prevalence	0	0	30%
PVO	2	1	23%

Figure 4-2: Comparative Summary for AQ15 and PVO on [Michalski, Mozetic, Hong, and Lavrac 86] Data

Method	Variables	Rules	Errors (set 1)	Errors (set 2)
C4 pruned rules	8	2	31	43
PVO random resampling	8	2	17	30

Figure 4-3: Comparative Summary for C4 and PVO on [Quinlan 87b] Data

One of the future goals of our research is to integrate the PVO procedure into a general rule refinement system.

References

[Michalski, Mozetic, Hong, and Lavrac 86]
Michalski, R., Mozetic, I., Hong, J., and Lavrac, N.
The Multi-purpose Incremental Learning System AQ15 and its Testing Application to Three Medical Domains.
In *Proceedings of the Fifth Annual National Conference on Artificial Intelligence*, pages 1041-1045. Philadelphia, Pa., 1986.

[Quinlan 87a] Quinlan, J.
Generating Production Rules from Decision Trees.
In *Proceedings of the Tenth International Joint Conference on Artificial Intelligence*, pages 304-307. Milan, Italy, 1987.

[Quinlan 87b] Quinlan, J.
Simplifying Decision Trees.
Int. Journal of Man-Machine Studies :in press, 1987.
also Tech. Report 87.4, New South Wales Intitute of Tecnology, School of Computing Sciences.

[Weiss, Galen, and Tadepalli 87]
Weiss, S., Galen, R., and Tadepalli, P.
Optimizing the Predictive Value of Diagnostic Decision Rules.
In *Proceedings of the Sixth Annual National Conference on Artificial Intelligence.* Seattle, Washington, 1987.
in press.

END
DATE
FILMED
6-1988
DTIC